THIRTEEN

Human Factors Problems in the Design and Evaluation of Key-entry Devices for the Japanese Language

CHARLES R. BROWN

Human factors studies on the design of keyboards are older than the field of human factors engineering itself. Most people generally associate the beginnings of the field of human factors engineering with the 1940s, a period that is some considerable time after pioneering studies of typewriter keyboards by Klockenberg in the 1920s (Kroemer 1972) and by Dvorak and his associates in the 1930s (Dvorak, Merrick, Dealey, & Ford 1936). Following these initial investigations, studies of keyboards, keysets, and data-entry devices have appeared in great numbers. Indeed, Seibel's review of the field (1972) contains no less than eighty relevant citations. Keyboards are the interface between man and a variety of very common machine devices in our society: telephones, typewriters, business machines, and computers of all kinds. Almost without exception, all of the research on the design of keyboards has been done with English-speaking persons using the English language.

Late in 1970, IBM Japan tested two Japanese language key-entry devices using procedures that I developed at IBM's Mohansic Laboratory in Yorktown Heights, New York. These tests brought us face to face with some human factors problems of much greater complexity than would be encountered in comparable work in any European country. At the same time, the complexity of the problem led me to a design principle that I think may have some general applicability.

Charles R. Brown, International Business Machines Corporation, Advanced Systems Development Division, 2651 Strang Blvd., Yorktown Heights, N.Y. 10598, U.S.A.

THE PROBLEM

By far the biggest problem in designing a Japanese keyboard is that it has to accommodate the Japanese language. The basic written language of Japan, called Kanji, was adopted from the Chinese approximately 1700 years ago. It consists of a large, indeterminate number of pictographic characters. For most purposes, however, 2,000 to 4,000 Kanji characters generally suffice. In addition to the Kanji characters, several other classes of character have to be included on a Japanese keyboard. These are:

1. Hiragana, a phonetic subset of about 50 characters used in Japanese syntax
2. Katakana, a phonetic subset of about 50 characters used to transcribe foreign expressions
3. Japanese numerals, the equivalent of 0 through 9 plus a few for large quantities such as 100 and 1,000
4. Arabic numerals, 10 in all
5. The Roman alphabet, consisting of 26 characters
6. Miscellaneous characters consisting of such things as TV channel numbers, baseball symbols, and special numerals for financial transactions.

Imagine trying to accommodate all of these character sets with conventional European typewriter keys. With one character per key, the keyboard would be about 1 meter wide and 1 meter high. Nevertheless, the keyboard can be a small fraction of this size without degrading the legibility of the character legends. However, approaching this theoretical minimum size requires keying techniques very different from those of the European typewriters. Our test keyboards illustrate this point.

THE TEST KEYBOARDS

One keyboard requires a stylus to actuate a key. The dimensions of the individual keys (or character positions) are somewhat over 4 mm, too small to depress with a finger. However, this size is quite adequate if the key is depressed with a pencil-like stylus. A very different keying technique for approaching minimum keyboard size is used with the second keyboard. On each key is a legend of 12 characters. These characters are arranged in an array of 4 rows and 3 columns. To key a character requires the use of two hands. The right hand depresses the key containing the desired character. The left hand selects the desired character from the 12 on the key by depressing one of 12 shift keys arranged in the same pattern as the legend characters. Because the shift keys are arranged in 3 columns, the little finger of the left hand is not used for shift-key operation.

KEYING TIMES

Naturally, we were greatly interested in the performance, that is, the keying times and error rates, associated with these two keyboards. Would the greater complexity of these devices, in comparison with conventional American and European devices, make a difference in performance? I shall use the concept of a word for purposes of comparison. Two characters, I am told, are roughly equivalent to one word in Japanese. Five characters are usually taken as the average word length in English. Using the card punch as a standard for keying English words, Klemmer and Lockhead (1960, 1962) report performance values for trained operators of:

- Keying times: 0.29 sec/character, or 1.45 sec/word
- Error rates: 0.05% characters in error, or 0.25% errors/word.

Comparable performance values for the Japanese shift keyboard in the application environment are:

- Keying times: 0.75 sec/character, or 1.5 sec/word
- Error rates: 0.2% characters in error, or 0.4% errors/word

Clearly, on a word or informational basis the speeds and error rates in both English and Japanese are similar. From this we conclude that great cultural variation has produced little or no variation in the level of skilled performance.

There are no comparable data on the use of the stylus keyboard in the application environment. However, let us take a look at what was learned about both keyboards in the test environment and how well the test results correlate with the results found in the application environment.

LAYOUT OF THE CHARACTERS

In the tests of the two Japanese key-entry devices, we decided to use the layout of the shift keyboard on both devices. The stylus keyboard had not yet been used in an application environment,[1] whereas the shift keyboard, called the Kantele (Kanji teletype), had been used for over twenty years in the newspaper industry. Fortunately, the Kantele layout could be placed on the stylus keyboard's physical configuration.

Figure 1 shows that the keys on the Kantele keyboard are arranged in quadrants consisting of 4 rows and 12 columns. Each key contains a 4- by 3-character array. The 12 shift keys are located in the lower left-hand corner. In this same general area are also some function keys that were not

[1]The stylus keyboard, an experimental device, was designed and developed by IBM's Advanced Systems Development Division in Yorktown Heights, New York (Juliusburger, Krakinowski, & Stilwell 1970), and was demonstrated at the World's Fair in Osaka, Japan, in 1970.

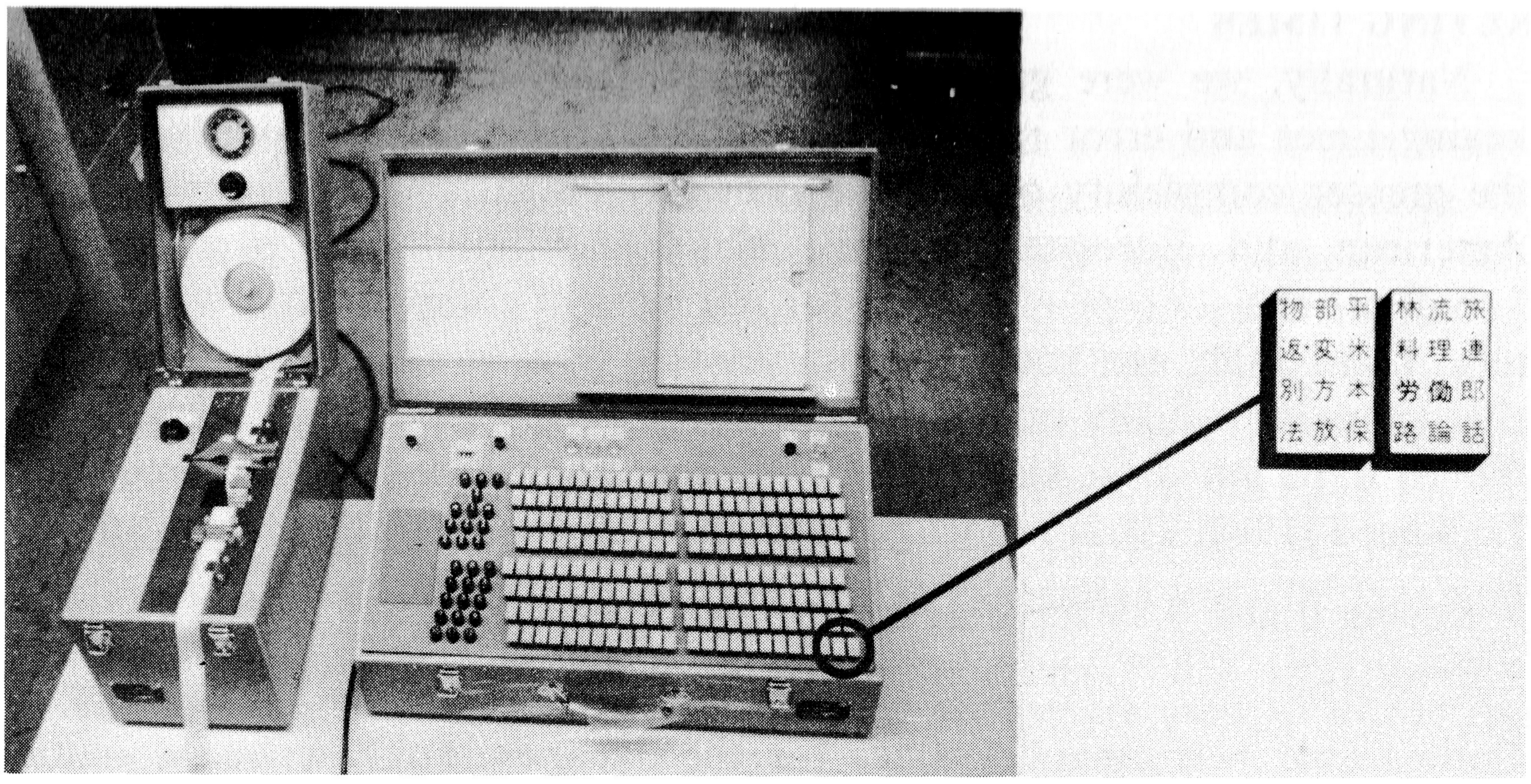
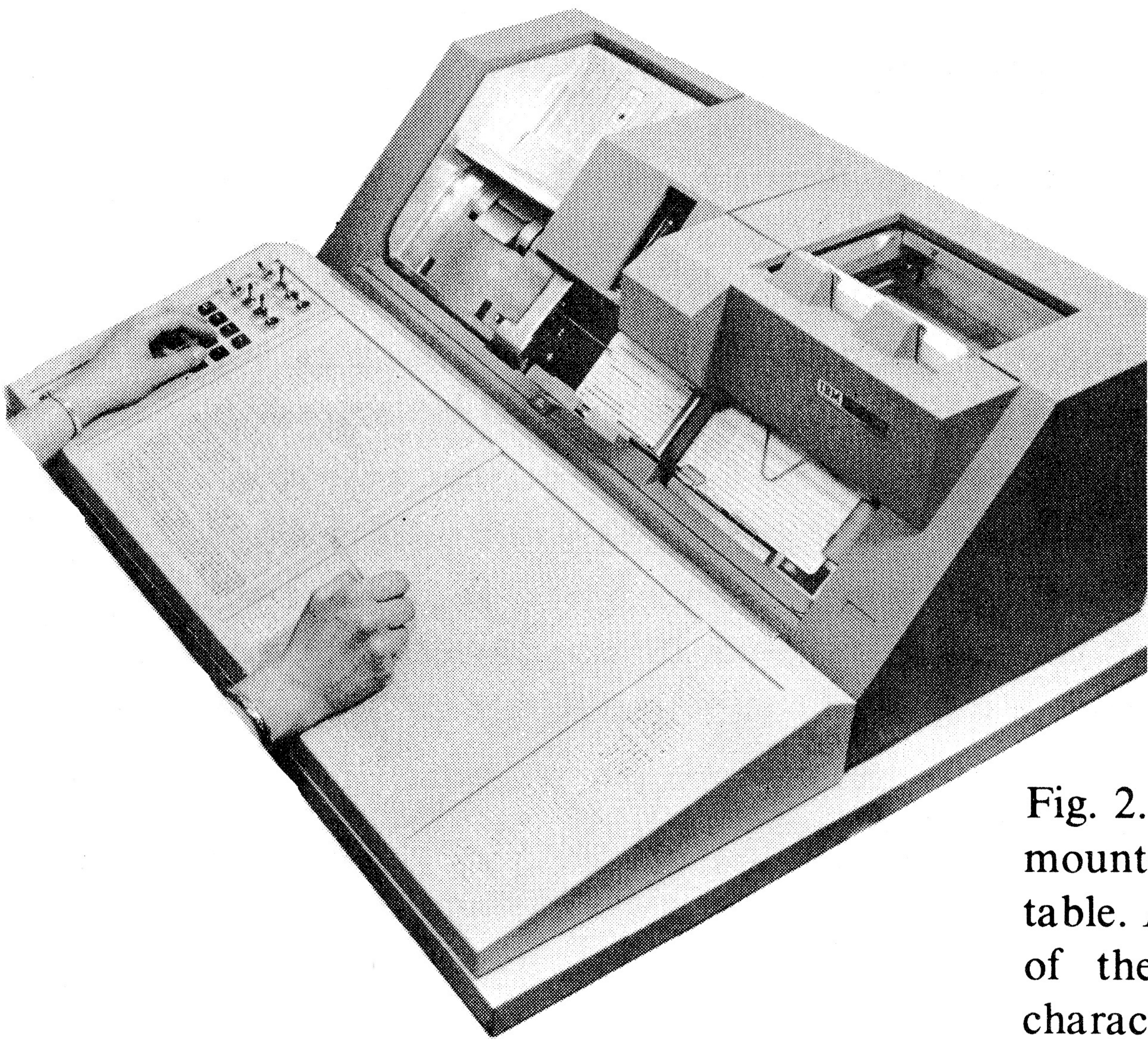


Fig. 1. The shift (Kantele) keyboard and paper tape unit.

used in the present study. Figure 2 shows the stylus keyboard configuration. Our interest is limited to the 2 large modules in the center. Each module has 8 single-character key arrays of 5 rows and 40 columns. A character key of the stylus keyboard is not similar to a conventional, mechanical key. A nomenclature sheet covers the switch positions. A character is keyed by depressing this nomenclature sheet approximately 1 mm with a stylus. Auditory feedback is provided by the cardpunch.

The result of placing the characters of a shift keyboard quadrant on four of the character arrays of the stylus keyboard is illustrated in Figure 3.



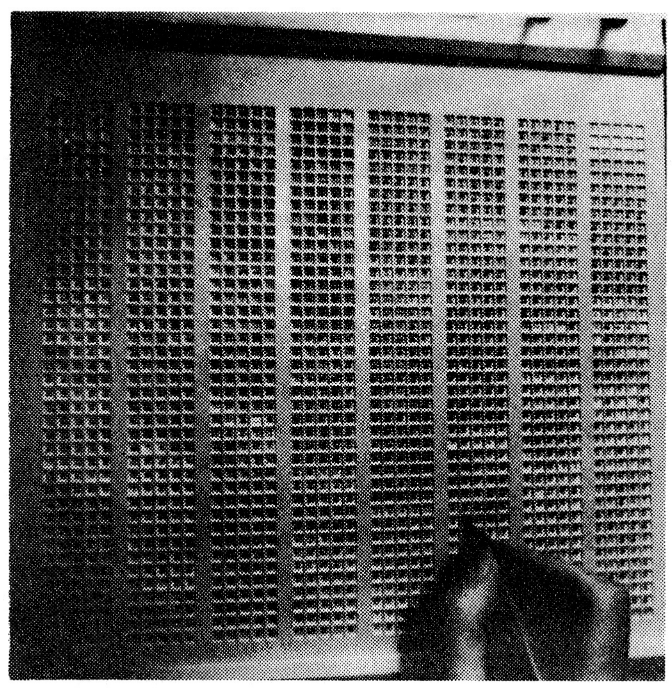
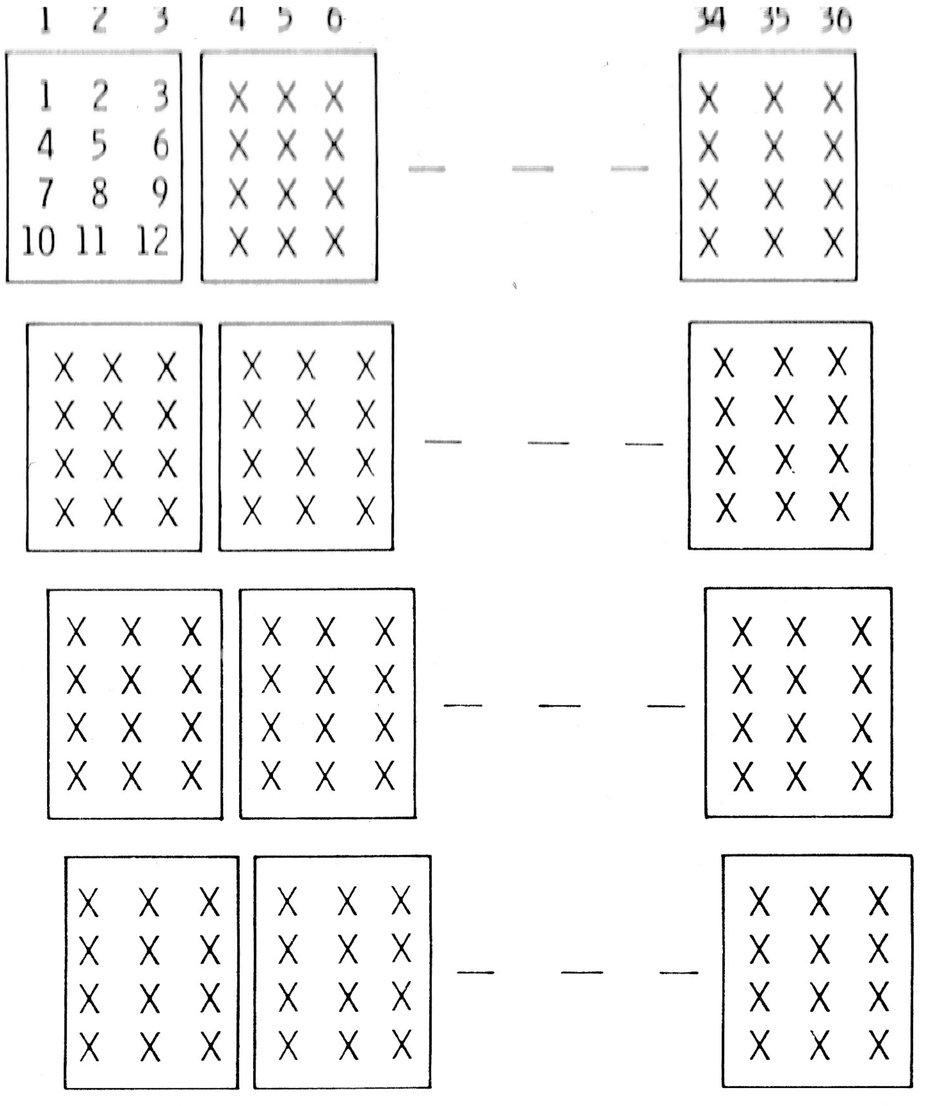
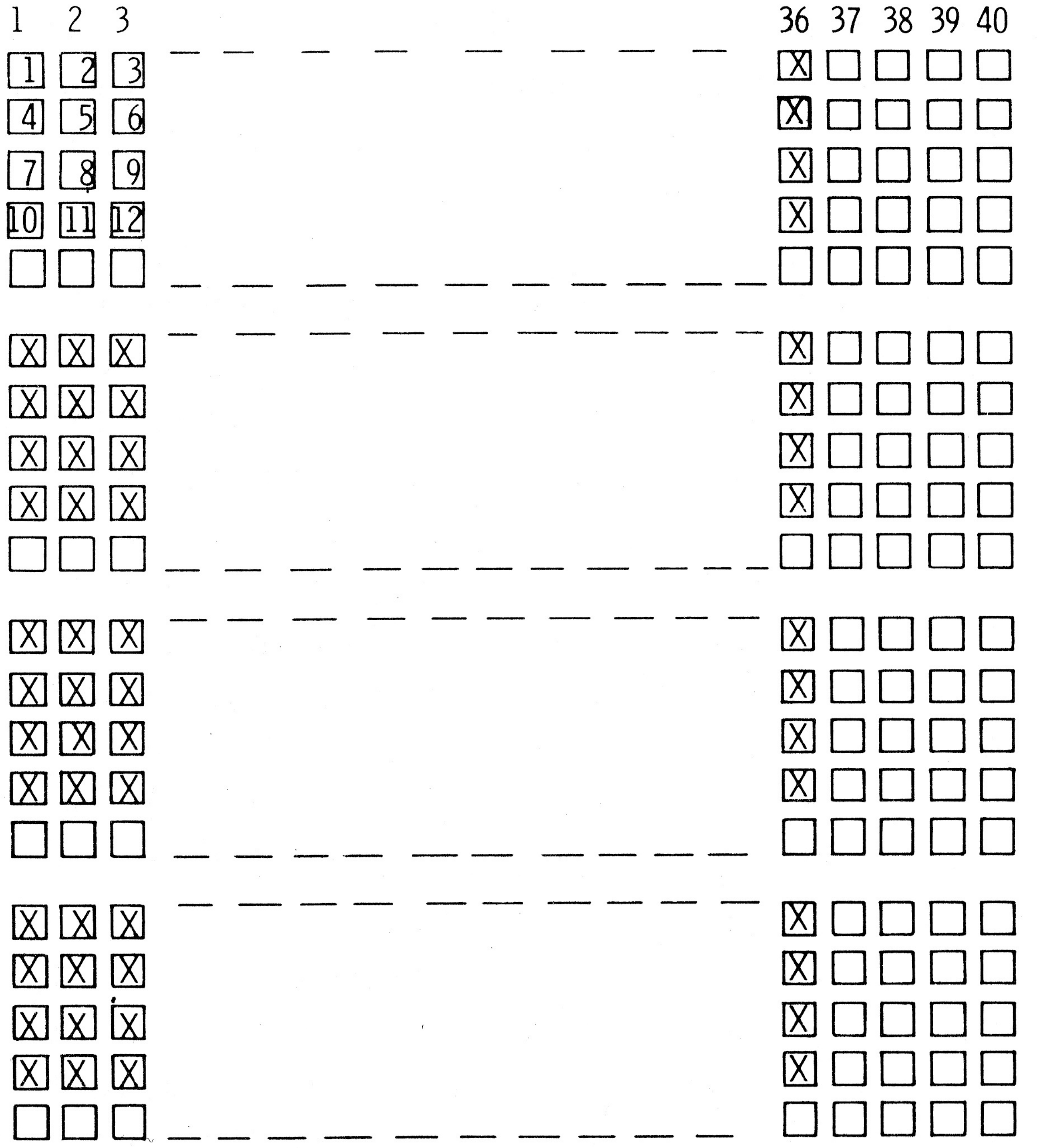


Fig. 2. *Left*: The stylus keyboard mounted on an IBM 029 cardpunch table. *Right*: A close-up view of one side of the stylus keyboard showing the character (key) positions and sizes.



SHIFT KEYBOARD QUADRANT



STYLUS KEYBOARD QUADRANT

Fig. 3. The shift keyboard layout placed upon the stylus keyboard configuration—one quadrant.

Placing one Kantele row on one stylus character array maintains the vertical perceptual grouping. Appropriate color coding of the stylus keyboard maintains the horizontal grouping of the separate keys of the Kantele. If the shift keyboard layout is placed on the stylus keyboard as indicated, a number of stylus keyboard key positions are not used. The fifth row of each character array is empty. Also, the right-most four columns of each character array are empty. A Kantele quadrant is 181 mm wide and 95.3 mm high. Placed upon the stylus keyboard, a quadrant increases in size to 217.8 mm wide and 135.8 mm high. The somewhat greater width of the stylus keyboard qudrant is due to the somewhat larger character area. The appreciably greater height of the stylus keyboard is due to the larger character area, the unused character rows, and the separation between character arrays. On both keyboards we are concerned with a total ensemble of 2,304 characters.

OPTIMUM LAYOUT

To get a fair test of these two physically dissimilar devices, we needed to assure ourselves that we had nearly optimum layouts. Performance by highly skilled operators on the Kantele keyboard in the application environment suggests that its layout is nearly optimum. Text is keyed at 1.00 sec/character after a few months and at 0.75 sec/character after 1 year. After 1 year, only 0.2 percent of strokes are in error. As previously noted, this is roughly equivalent to the performance of an operator trained on a conventional card punch in the United States.

An analysis of the Kantele's layout into quadrants supported our feeling that the Kantele layout is close to optimum. The primary organization is by usage. The characters with the highest usage are located in the lower right-hand quadrant. Usage decreases from quadrant to quadrant in a counter-clockwise fashion. The human factors benefit from this organization is that the average area searched and the average distance reached are greatly reduced, so that the effective size of the keyboard is greatly reduced. The high-usage region contains far less than half the characters, but these characters are used far more than half the time. The odds are high that a character picked at random will be in the high-usage region. Figure 4 shows the cumulative usage accounted for by Kanji characters rank-ordered from the most used to the least used. The 200 most used Kanji characters account for over 50 percent of the usage.

To pursue the benefits of organizing a keyboard by usage, consider first a layout in which the entire keyboard (containing all the Kanji characters) is one region. Now rearrange the characters into two separate regions, one containing the characters used more often and the other the characters used less often. Since search and reach in the high-usage region of the two-region keyboard are less than those in the one-region keyboard, the rate of keying within the former will be improved over the keying rate that existed

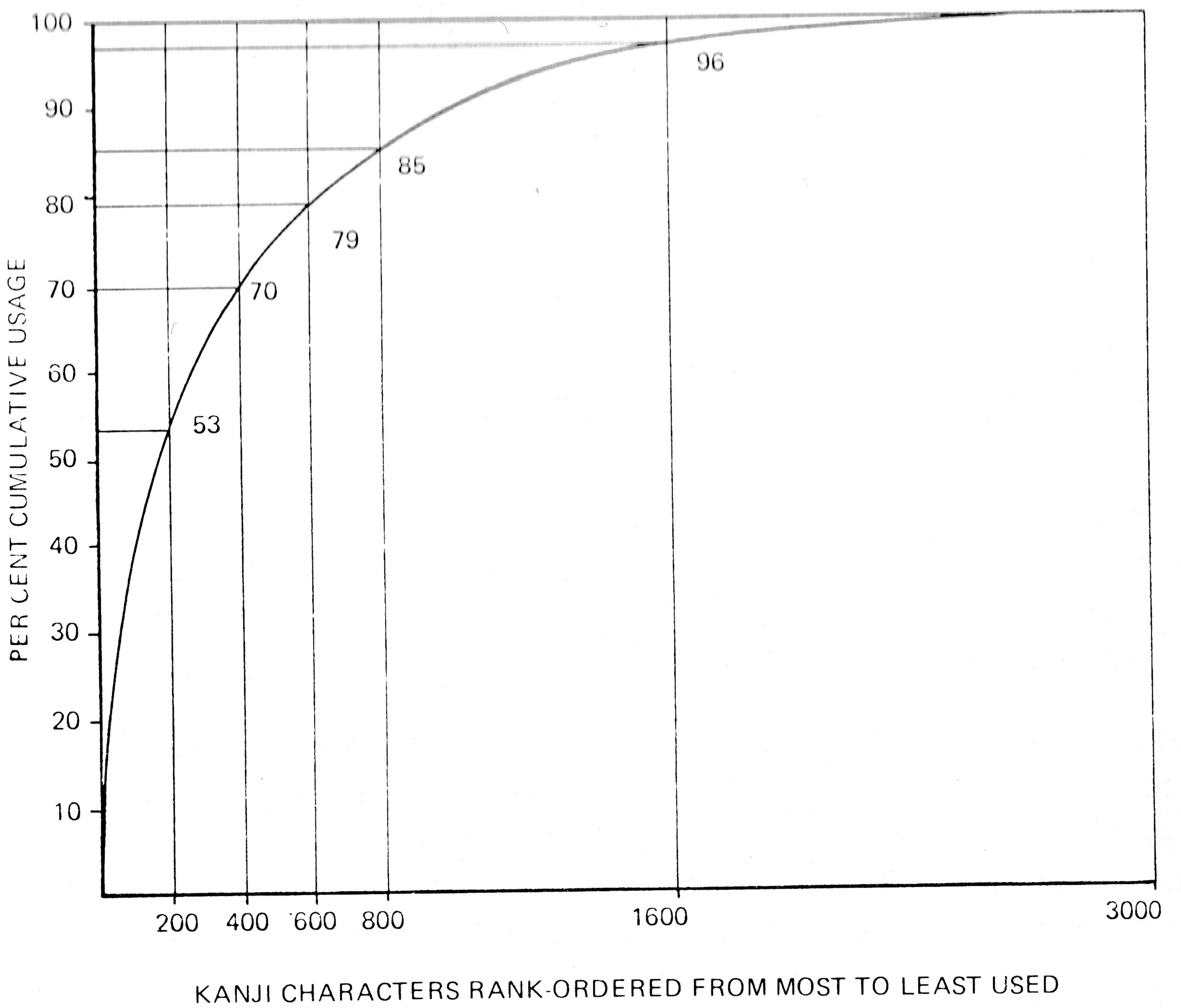


Fig. 4. Percent of usage accounted for by Kanji characters rank-ordered from most used to least used.

before the keyboard was partitioned. This saving in the time to key a character within the high-usage region is amplified by the odds that a character is in that region to produce a gain in time available to key within the low-usage region. This gain results in an improved keying rate for the keyboard as a whole, since no additional time is actually required to key a character within the low-usage region.

SIZE OF THE HIGH-USAGE REGION

How large should the high-usage region be? The answer involves a trade-off between savings and odds. With a small high-usage region, the odds that a character will be in that region may be so negligible that it would attenuate appreciable savings in keying speed. With a large high-usage region, the savings in keying speed may be so negligible as to attenuate appreciable odds. Thus, the layout design problem is to estimate a size of high-usage region that will have *moderate odds together with moderate savings in keying speed.* The Kantele layout appears to be rather close to the optimum solution to this design problem.

Quantitatively, these ideas take the following form. Let the average time to key a character before the keyboard is partitioned into a high-usage region, R_1, and a low-usage region, R_2, be K. Maintaining the overall

keying time, we assume that the time to key a character within R_1 is reduced by an amount Δ_1. This results in an excess time, Δ_2, to key a character in R_2. Therefore, the average time (K) to key a character on the entire keyboard is expressed by:

$$P_1 (K - \Delta_1) + (1 - P_1)(K + \Delta_2) = K, \tag{1}$$

where P_1 is the probability of keying in R_1. Solving for the gain, Δ_2, in terms of savings, Δ_1, we have:

$$\Delta_2 = \left(\frac{P_1}{1 - P_1}\right) \Delta_1 . \tag{2}$$

Equation (2) says that the gain is the savings amplified by the odds of keying in the high-usage region. For example, if the savings were 1.0 sec and the odds were 9 to 1 ($P_1 = 0.9$), the gain would be 9.0 sec.

However, since K is sufficient time to key anywhere on the unpartitioned keyboard, it is sufficient time to key a character within R_2, given that an operator has learned whether or not a character is in R_1. Thus, we can calculate the minimum improvement in keying time to be expected over the entire keyboard by setting equation (1) equal to K_n, the new keying time, and Δ_2 equal to 0. This gives us:

$$\frac{K_n}{K} = \frac{K - P_1\Delta_1}{K} . \tag{3}$$

This expression shows that as the usage in R_1 approaches unity, the average time to key a character for the keyboard as a whole approaches the time to key a character in R_1. Figure 5 illustrates this graphically. The axes may be interpreted as follows:

K_n/K = Either (a) the average time (sec/character) to key a character on the partitioned keyboard relative to the average time to key a character on the unpartitioned keyboard or (b) the keying rate (characters/sec) for the unpartitioned keyboard relative to the keying rate for the partitioned keyboard. As this number decreases from 1.0, keying on the partitioned keyboard improves relative to keying on the unpartitioned keyboard.

$(K - \Delta_l)/K$ = Either (a) the average time (sec/character) to key a character within the high-usage region relative to the average time to key a character for the unpartitioned keyboard or (b) the keying rate (characters/sec) for the unpartitioned keyboard relative to the keying rate within the high-usage region. As this number decreases from 1.0, keying in the high-usage region improves relative to keying on the unpartitioned keyboard.

Recall that to achieve a $P_l = 0.9$ or so requires relatively few of the characters to be in R_l. Approximately 800 Kanji and 50 Hiragana characters account for 90 percent of the usage for standard text. The keying time in a high-usage region consisting of these characters would essentially determine the average keying time for a keyboard in excess of 3,000 characters. So the high-usage region need not exceed approximately 25 percent of the keyboard. But how small can the high-usage region be? Note that if $P_l = 0.5$, equation (3) shows that the keying within R_l must take no time at all if the over-all average is to be 0.5 of K (that is, $\Delta_l = K$). Thus, R_l must contain a large enough set of high-usage characters so that the prob-

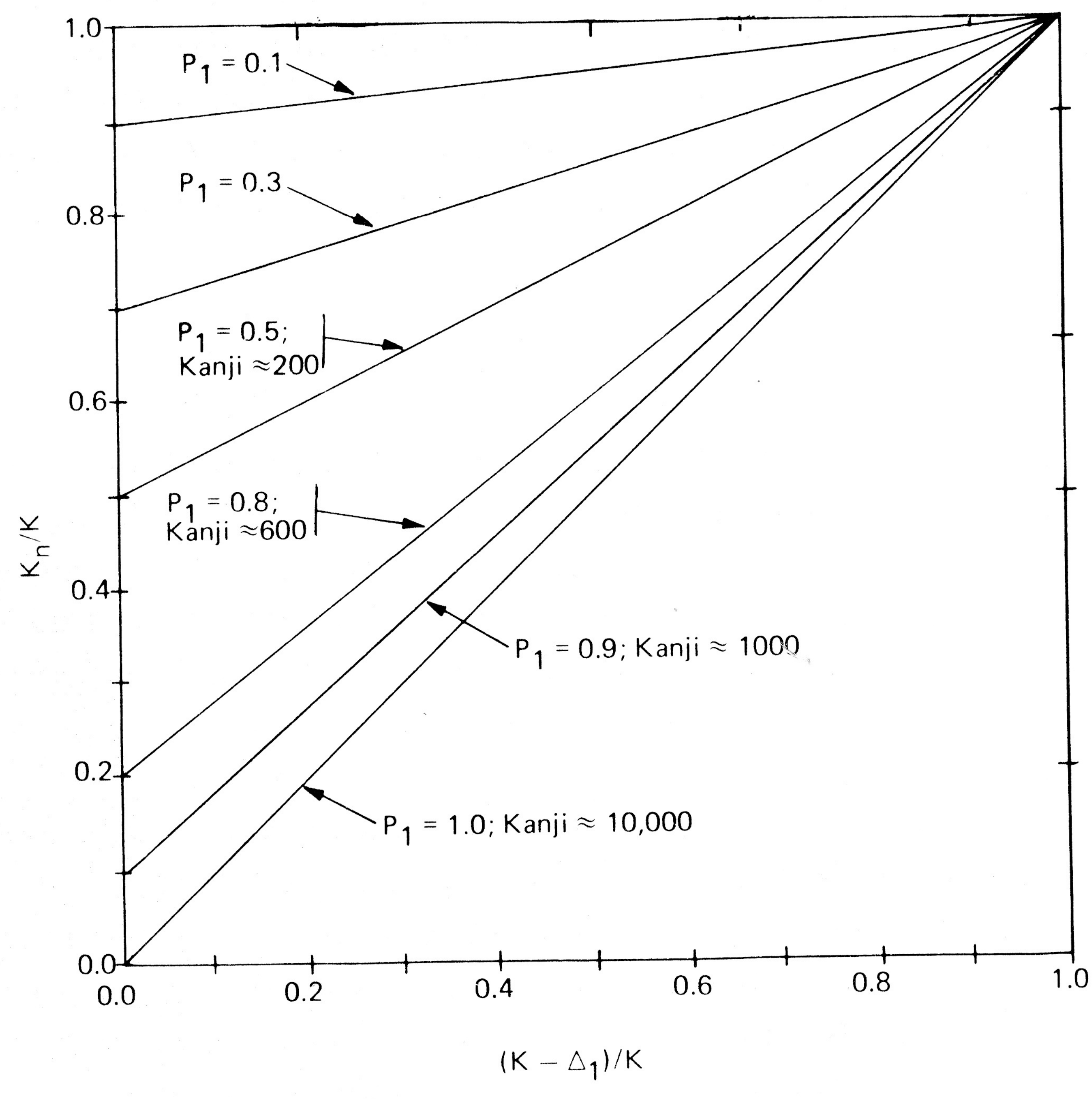


Fig. 5. Improvement of keying throughout the total keyboard as a function of the improvement of keying in the high-usage region.

ability of keying in the area is substantially greater than 0.5. The Kanji usage curve (Fig. 4) in conjunction with Figure 5 locates a relatively narrow range within which the optimum lies. My judgment is that an R_1 within the range from approximately 600 Kanji characters ($P_1 \cong 0.8$) to 1,000 Kanji characters ($P_1 \cong 0.9$) is nearly optimum.

CHARACTER-SEQUENCE RULE

Also important is the organization of characters within a keyboard region. On the keyboards studied, the Kanji characters within a region are arranged by a pronunciation sequence, the Hiragana. The first syllable of each character's name is used as an index of arrangement. The secondary rule of arrangement is the number of strokes that make up the character. Thus, a dozen or more Kanji having the same first syllable may be grouped together.

The Kanji characters are sequenced in columns. A column is 3 characters wide (one Kantele key) and 16 characters high (four Kantele keys). The order of characters within a column is from left to right across the 3 characters in a line and from top to bottom through the lines of 3 characters. Succeeding columns are to the right. On the stylus keyboard, adjacent columns were of different colors as a perceptual aid. Exceptions to this scan rule occur when other character subsets are within a quadrant. For example, the Hiragana subset is located centrally within the high-usage quadrant. This permits 80 percent to 90 percent of standard text to be keyed from this quadrant.

EVALUATION PROCEDURES

For reasons of time and cost, the evaluation had to be conducted with one keyboard of each type. To approach the level of training of full-time operators in the application environment, each operator in the evaluation was trained on a mock-up and tested on the associated keyboard. The mock-ups were essentially just layouts of the character positions. The similarity between mock-up and keyboard was high for the stylus keyboard, since key depression on the keyboard is minimal. The similarity between the mock-up and the shift keyboard was low, except for character position. Shift-key training was limited to locating the shift-key position by touch.

Training sessions were of two types: general practice, in which regions of the keyboard were studied for periods of time proportional to the average usage of characters in the region, and document practice, in which four 40-character documents were rehearsed for 50 minutes. Both types of practice sessions occurred in the morning and afternoon.

Testing also occurred both in the morning and afternoon. The four documents rehearsed in the preceding practice session were timed. Two of these documents were segments of text, a mixture of pictographic Kanji and

phonetic Hiragana. One document was text with Kanji deleted—pure Hiragana. One document was text with Hiragana deleted—pure Kanji. A similar set of four unrehearsed documents was timed during test sessions. These new documents became the documents for the next practice session. Unrehearsed text was not tested until the second week, when layout familiarity permitted its inclusion within the scheduled time period. The operators had immediate knowledge of their time scores. Information on their error scores was delayed by two or three days.

Ten weeks of evaluation were planned. However, one week of data collection was lost in general orientation procedures. Five girls were hired for each keyboard. We lost one operator from each group: one resigned, and the other was absent so much that her data could not be used. The groups were matched according to a standard card-punch operator examination.

RESULTS

UNREHEARSED TEXT

For keying times no statistically significant difference was obtained between the learning curves for the two keyboards. At week 2, keying was somewhat under 3.50 sec/character; at week 9, the keying time was somewhat under 2.00 sec/character. When the test ended, the improvement curve for keying time had still not approached its asymptotic value. Yet the improvement in keying time was substantial, and the level finally reached appears to be fairly close to what probably occurs in the application environment.

The shift keyboard's error rate for unpracticed text was substantially higher than the asymptotic value of the application environment. This is reasonable. Since insufficient shift-key training preceded this evaluation, the test procedures did not simulate mastery. Throughout the test, the stylus keyboard's error rate for unpracticed text remained near the asymptotic value characteristic of the application environment.

PRACTICED TEXT

Keying time. Figure 6 shows the improvement in keying time for both keyboards for Hiragana, text, and Kanji. The most striking feature of these data is the rapid improvement during the first four weeks and the ensuing stability, as shown by the gradual improvement, after week 4. For the Hiragana and Kanji curves, this means relative stability after approximately eight hours of document practice. For the text curves, this means relative stability after approximately sixteen hours of document practice.

Another striking feature of these data is their regularity. First, the Hiragana, text, and Kanji curves for a particular keyboard never cross. For every week, the data are ordered by size of the character set: Hiragana

fastest, then text, and Kanji slowest. Second, corresponding curves for the two keyboards never cross. The stylus curve is *always* below its corresponding shift-keyboard curve. The third striking feature of these data is the rapid approach of the stylus keyboard's curve for text to the asymptote characteristic of the application environment. At week 4, keying with the stylus keyboard is at 0.99 sec/character. At week 9, keying is at 0.87 sec/character.

Last, but not least, note the fairly small increase in keying time with the increase in size of the character set. After week 4, for example, text is

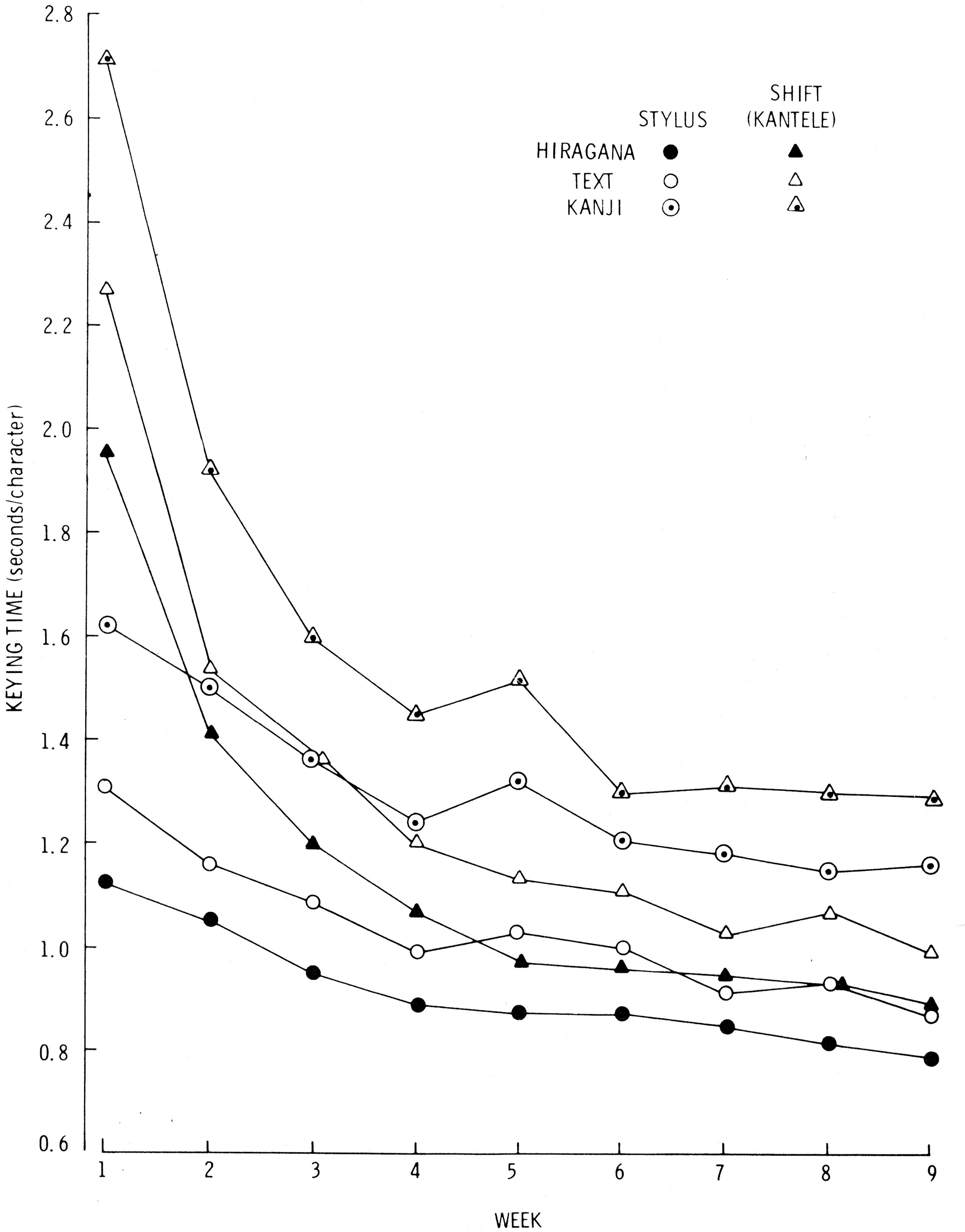


Fig. 6. Keying times for practiced material.

keyed about 0.10 sec/character slower than Hiragana. Figure 7 shows the keying time for text as a linear function of the keying time for Hiragana. Table 1 gives the results of the correlation and regression analyses on those keying times. The same linear function fits the data for both keyboards. Indeed, the fit is to within 1 percent of the total variance. The function through the Kanji keying times is drawn parallel to the curve through the text keying times to show the similarity between the two sets of data.

Error rate. Shift-key operation was not mastered during the course of the evaluation. Figure 8 shows erratic improvement for the shift keyboard

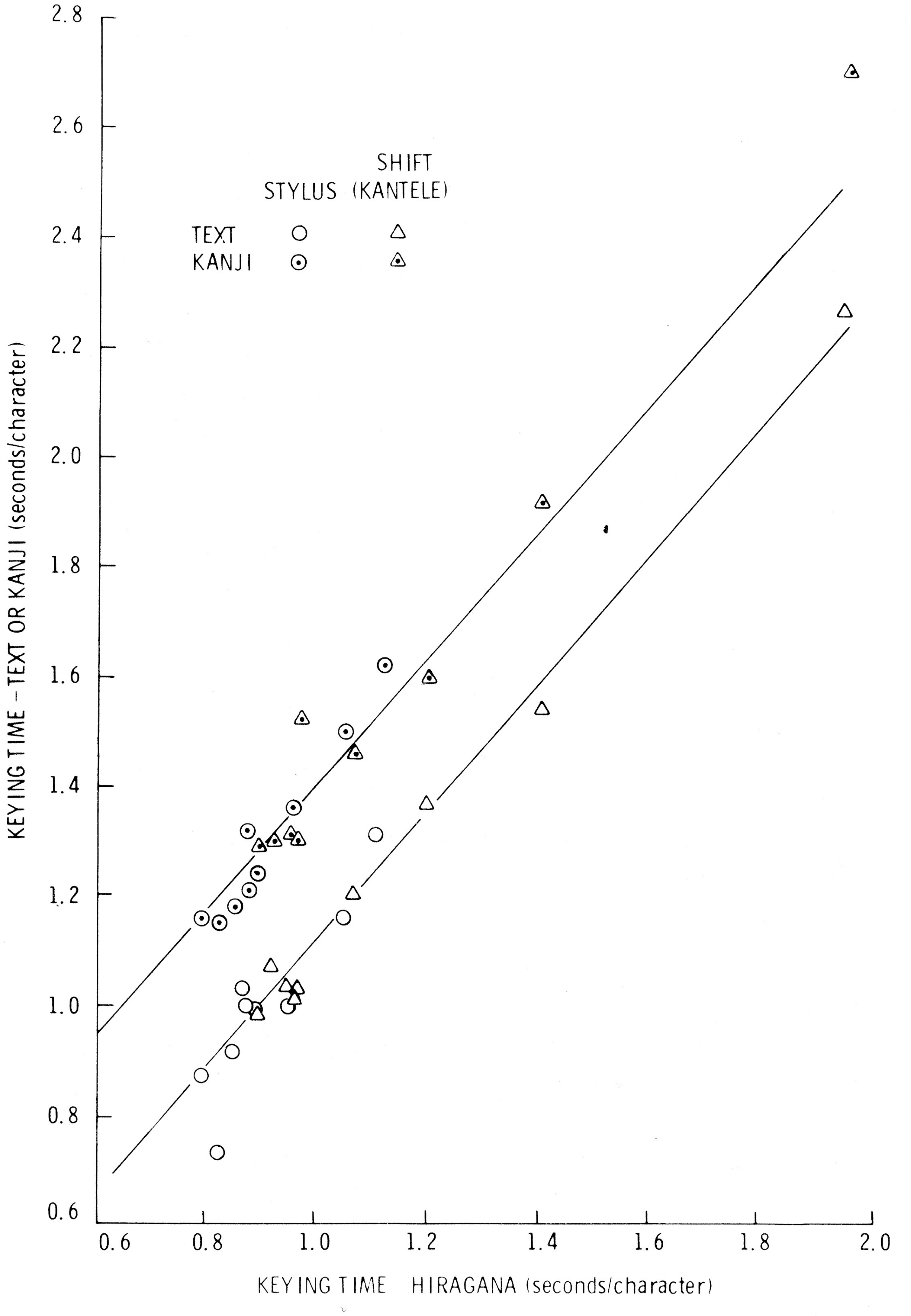


Fig. 7. Keying times for text and pure Kanji as a function of the keying time for pure Hiragana.

Table 1. Correlation and regression analysis for the data in Figure 7. The data for both keyboards have been combined

Variable			
X	Y	r^2	Regression equation
Hiragana	Text	0.9896	$Y = 1.16X - 0.03$
Hiragana	Kanji	0.9779	$Y = 1.32X + 0.09$

throughout these tests, and the final level is not near the asymptote usually found in the application environment. This is reasonable because the shift-key training that preceded the tests was lacking in realism, and the testing itself did not simulate mastery.

By contrast, the median error rates for the stylus keyboard were near the asymptote typical of the application environment (see Table 2). This is reasonable, since the stylus keyboard does not require learning the shift-key operation. Because the limiting factor in performance is the size of the character set and not the shift-key operation, in a sense the stylus keyboard simulates shift-key mastery.

A finding of some interest is that the error rate did not increase with the size of the character set for practiced material. Figure 9 shows text error predicted from Hiragana error. Table 3 shows that for both keyboards combined, errors in Hiragana can predict well over half (58 percent) of the variance in errors in text. The identity function ($Y = X$), is drawn to indicate that error rate is almost entirely insensitive to the size of the character set.

DISCUSSION

If I had not observed professional operators perform on the 12-shift keyboard, I probably would have overestimated the time required to key a character. The speed with which work can be done on such a keyboard is almost unbelievable. The operators are impervious to distractions and noise and are able to read, seemingly without difficulty, hand-written script with marks of correction. Skilled performance of such a complex task is truly impressive.

The excellent keying rate on the 12-shift keyboard is certainly not accidental. Mastering 12 shifts takes time, but it can be done. By exploiting the adaptability of the human, the size of what would otherwise have to be a very large keyboard has been greatly reduced. Over time, learning compensates for the complexity of the shift-key keyboard. However, the very large physical size of the keyboard is a permanent, limiting factor on performance.

The stylus keyboard is a solution to the problem of keyboard size that is quite different from the multiple-shift approach. It presents no insoluble

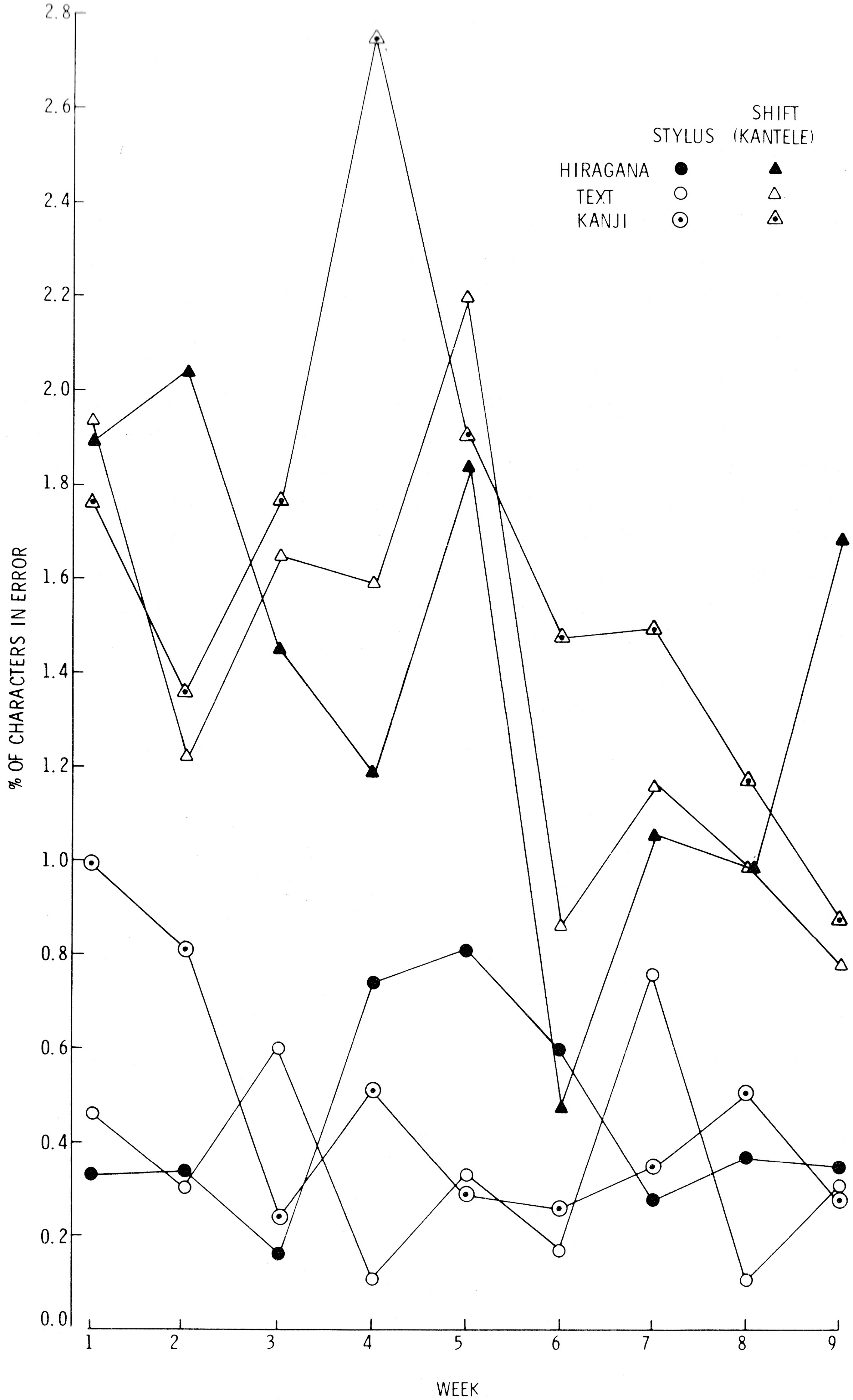


Fig. 8. Error rates for practiced material.

Table 2. Measures of central tendency of the percentages of characters in error for the stylus keyboard (see Fig. 8)

Test material	Mean	Approximate median
Hiragana	0.43	0.36
Text	0.36	0.25
Kanji (last 7 weeks)	0.26	0.25

problems and essentially eliminates the error-learning curve characteristic of shift-key operations. Furthermore, stylus operation permits practiced material to be keyed somewhat faster than does shift operation.

KEYBOARD ORGANIZATION

In addition to comparing the shift and stylus operations necessary to reduce the size of the keyboard, we have made use of a principle of keyboard organization by usage. The exploitation of this statistical property of the language markedly enhances the keying rate. Comparable examples of reducing the complexity of difficult tasks would be welcome additions to our reference works on human factors engineering.

Even though the layout of the keyboards into usage regions reduces the average distance between successive characters of text, we are still sur-

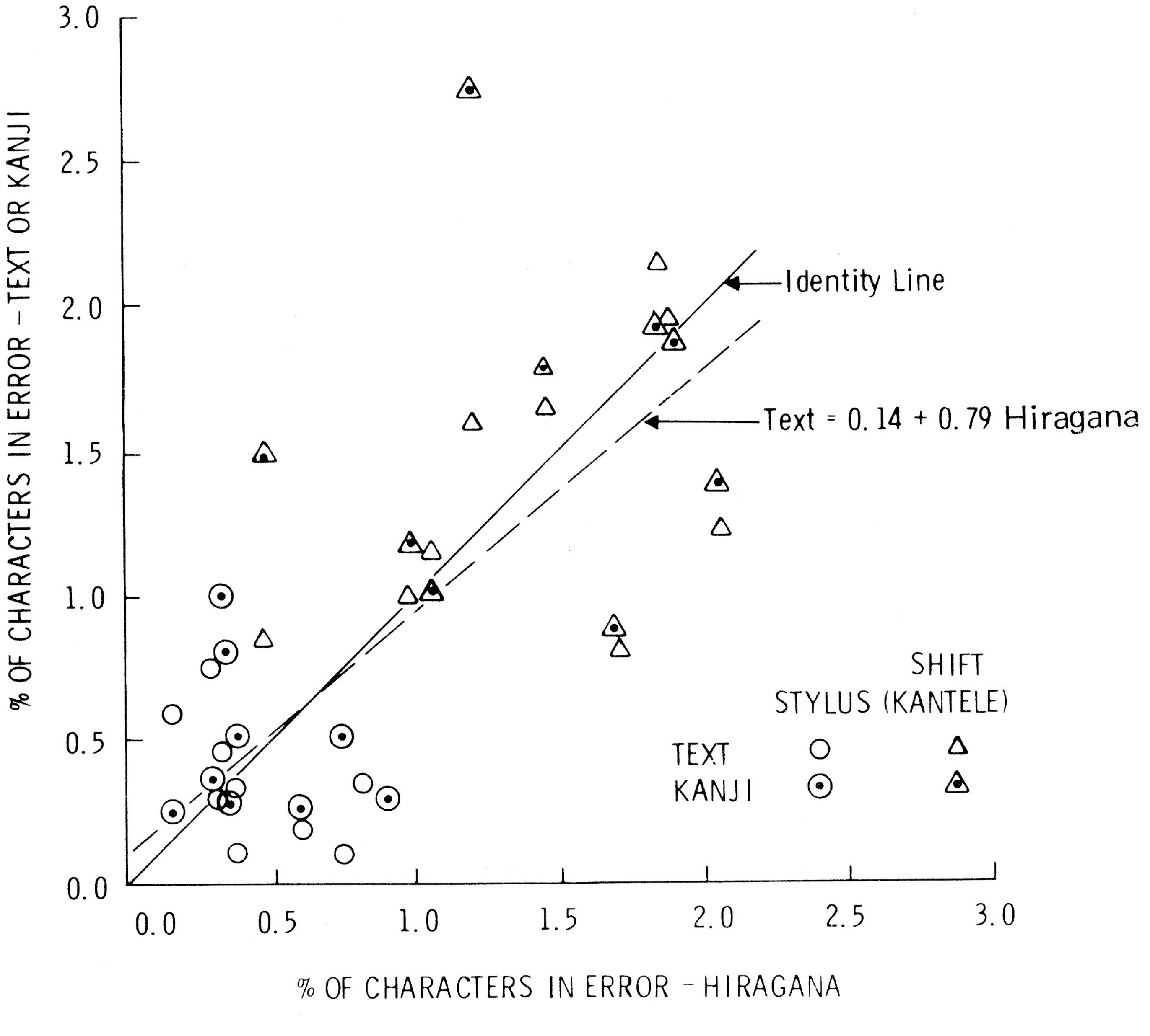


Fig. 9. Error rates for text and pure Kanji as a function of the error rate for pure Hiragana.

Table 3. Correlation and regression analysis for the data in Figure 9. The data for both keyboards have been combined

Variable			
X	Y	r^2	Regression equation
Hiragana	Text	0.583	$Y = 0.79X + 0.14$
Hiragana	Kanji	0.402	$Y = 0.73X + 0.37$
Hiragana	Text and Kanji	0.473	$Y = 0.76X + 0.26$

prised by the relative insensitivity of performance to the size of the character set once the layout by usage has been established. Hiragana is keyed about 10 percent faster than text, whereas Kanji is keyed about 20 percent slower than text. Clearly, our layout must be nearly optimum, for no layout would be expected to reduce the keying of text below that of a small 50-character subset, that is, the Hiragana. Thus, the size of the character-set, which determines the duration of learning, may have little effect upon the final level of skilled performance. It appears that the difficulty of a task or device is more appropriately measured by the time it takes to acquire the skill to perform the task, or operate the device, than by the level of asymptotic performance achieved on the task or device.

SHORTCUTTING TRAINING

Our attempt to simulate the performance of a skilled operator was successful. Our operators, after only a relatively few hours of document practice, exhibited the kind of stable keying performance one expects of truly skilled operators. The stylus keyboard operators approached very nearly the keying rate of operators in the application environment. This occurred even though our operators (a) were temporary employees rather than professionals, (b) were of a different sex from operators who normally do this kind of work, and (c) spent most of the day working with a mock-up that did nothing. Given the many differences between the test and application environments, our estimate of skilled performance is remarkably good.

The rapid improvement and ensuing stability of our data are encouraging. If we can show that other complex skills can be estimated with techniques similar to those used here, we will be able to evaluate the influence of various experimental conditions upon skills in 1 percent or so of the normal time required to acquire the skill.

SUMMARY

Japanese language keyboards may contain 100 times as many characters as European typewriters. If the keyboard is to approach the minimum size consistent with legibility requirements for the characters, keying must involve stylus operation or multiple shifts, or both. The average area

searched and the average distance reached can be reduced further by grouping the characters into two or more regions by usage. In that way, the odds can be increased that a character of text will be found in a relatively small, high-usage region. That in turn reduces the effective size of the keyboard. The use of either keying technique in conjunction with the layout principle results in performance comparable to that typical of skilled American card-punch operators. The enormous difference between the character sets and the keying techniques in the two cultures changes neither the average number of words keyed per unit of time nor the average number of errors made per word.

To estimate the performance level of fully trained operators, practiced material was used to simulate memory for character position. A few hours of such practice resulted in a level of performance that normally takes months to acquire. Thus, the asymptotic level of keying the Japanese language may be estimated without actually training operators to that level.

REFERENCES

Dvorak, A., Merrick, N. I., Dealey, W. L., and Ford, G. C. *Typewriting behavior.* New York: American Book Company, 1936.

Juliusburger, H. Y., Krakinowski, M., and Stilwell, G. R., Jr. The Kanji entry feature. Technical Report No. 17-245, IBM Advanced Systems Development Division, Yorktown Heights, New York, March 1970.

Klemmer, E. T., and Lockhead, G. R. An analysis of productivity and errors on keypunches and bank proof machines. Research Report RC-354, IBM Research Center, Yorktown Heights, New York, November 1960.

Klemmer, E. T., and Lockhead, G. R. Further data on card punch operator performance. Research Note NC 39, IBM Research Center, Yorktown Heights, New York, 1962.

Klemmer, E. T., and Lockhead, G. R. Productivity and errors in two keying tasks: A field study. *Journal of Applied Psychology*, 1962, **46**, 401–8.

Kroemer, K. H. E. Human engineering the keyboard. *Human Factors*, 1972, **14**, 51–63.

Seibel, R. Data entry. Chapter 7 in Van Cott, H. P., and Kincade, R. G. (Eds.) *Human engineering guide to equipment design* (rev. edition), Washington, D.C.: U.S. Government Printing Office; 1972.